DTU Library

Timbral aspects of reproduced sound in small rooms. I

Bech, Søren

Published in:
Acoustical Society of America. Journal

Link to article, DOI:
10.1121/1.413047

Publication date:
1995

Document Version
Publisher's PDF, also known as Version of record

Timbral aspects of reproduced sound in small rooms. I

Søren Bech[a)]

The Acoustics Laboratory, Building 352, Technical University of Denmark, DK-2800 Lyngby, Denmark

(Received 11 November 1992; revised 31 May 1994; accepted 2 October 1994)

This paper reports some of the influences of individual reflections on the timbre of reproduced sound. A single loudspeaker with frequency-independent directivity characteristics, positioned in a listening room of normal size with frequency-independent absorption coefficients of the room surfaces, has been simulated using an electroacoustic setup. The model included the direct sound, 17 individual reflections, and the reverberant field. The threshold of detection and just-noticeable differences for an increase in level were measured for individual reflections using eight subjects for noise and speech. The results have shown that the first-order floor and ceiling reflections are likely to individually contribute to the timbre of reproduced speech. For a noise signal, additional reflections from the left sidewall will contribute individually. The level of the reverberant field has been found to have an effect on the contribution of the individual reflections. An increase in the level of individual reflections are most likely to be audible for the first-order floor and ceiling reflections, and certain reflections from the sidewalls.

PACS numbers: 43.55.Hy, 43.55.Jz, 43.66.Jh

INTRODUCTION

Although the perception of sound in rooms has been of interest to loudspeaker manufacturers for many years, the interaction between a loudspeaker, the listening room, and the listener is still not well understood. Therefore, it is very difficult for a manufacturer to predict the quality of the sound produced by a loudspeaker in a specific environment.

An obvious target for the design of an improved loudspeaker is one where reproduction is virtually independent of the listening room, and the position of the loudspeaker in the room. This could be possible through the application of sophisticated methods for control of the acoustic properties (e.g., directivity and frequency response) of loudspeakers, if the perception of reproduced sound was better understood. With this purpose in mind, a five-year research project, "Archimedes," was started in 1987 as a joint project among Bang and Olufsen A/S (DK), KEF Audio Ltd. (GB), and The Acoustics Laboratory of the Technical University of Denmark. The project has been financed under the European research program, EUREKA.

It was decided to concentrate experimental investigations on the timbre of reproduced sound as perceived by a listener in a typical room. To prevent confusion caused by simultaneous perception of timbre and localization, it was decided to investigate a monophonic reproduction system, represented by the right-hand loudspeaker of a stereophonic setup.

Two basic questions have been investigated:

(1) Which early reflections are sufficiently strong to contribute individually to overall timbre, and which only contribute collectively?

(2) How much must the level of an individual reflection change to produce a change in the overall timbre of the sound field?

Both questions are related to the directivity of a loudspeaker. The first is related to the basic design of a loudspeaker with fixed directivity, and the second to a loudspeaker that allows control of the directivity.

The influence of individual single reflections on the timbre of the sound field has been investigated previously. Atal *et al.* (1962), Sommerville *et al.* (1966), Bilsen (1968), Zurek (1979), Olive and Toole (1989), and Bech (1989) have all established threshold values for single reflections in combination with the direct sound as a function of the delay of a reflection. However, only Olive and Toole (1989) used a total sound field that included more than the direct sound, the investigated reflection, and perhaps one or two additional reflections. Olive and Toole (1989) investigated the threshold for a single lateral reflection in combination with the direct sound in a real room, including its reflections. The present investigation examines several individual reflections with the correct relationship between angle of incidence and time delay relative to the direct sound, in the presence of multiple other controllable reflections and a reverberant field.

The paper is organized as follows: Secs. I–IV contain a description of the setup, stimuli, subjects, and the general procedure. Section V contains the results and discussion, Sec. VI a general discussion, and Sec. VII a summary of the findings.

I. EXPERIMENTAL SETUP

An examination of the influence of individual parts of the sound field (e.g., reflections) necessitates the use of a simulation technique. Several techniques were examined and it was decided that electroacoustic simulation offered the best possibilities.

Electroacoustic simulation of a sound field requires a listener surrounded by loudspeakers in an anechoic chamber.

[a)]Present address: Bang and Olufsen A/S, Peter Bangs Vej 15, DK-7600 Struer, Denmark.

 0001-4966/95/97(3)/1717/10/$6.00

Reichardt and Lehmann (1978) have shown that a simulated sound field should include (1) the direct sound, (2) discrete early reflections, and (3) a subjectively diffuse sound field in order to convey a sufficient degree of room impression to the listener.

The direct sound and early reflections are usually represented individually by single loudspeakers, and the reverberant field by a group of loudspeakers. For practical and acoustical reasons (to keep the room anechoic) the number of loudspeakers in the setup must be limited. The direct sound requires one loudspeaker, and experience has shown that the subjectively diffuse field requires a minimum of four to six loudspeakers. The total number of loudspeakers is thus determined by the number of individual reflections that need to be modeled. Therefore a choice has to be made as to the time of the impulse response at which discrete reflections give way to a subjectively diffuse sound field.

The early part of the sound field is characterized by an increase in impulse density as a function of time. However, it is well known that the human hearing system, shortly after onset of a continuous sound, perceives the sound field as a continuous whole. This suggests that the hearing system only follows changes in impulse density to a certain point, and thereafter the sound field is perceived as continuous, irrespective of a continuing increase in impulse density.

Schreiber (1960) investigated this hypothesis and found that subjects (50%) cannot detect an increase in impulse density if the mean impulse density is higher than 2000 s^{-1}. This applies for low-pass filtered signals, with cutoff frequencies of 2 or 4 kHz. Thus it can be assumed that a sound field will be perceived as continuous if the mean impulse density is higher than 2000 s^{-1}.

The mean impulse density as a function of time for a rectangular room with reflecting walls can be approximated by

$$\frac{\Delta N}{\Delta t}=\frac{4\pi c^3}{V}t^2 \ \ (\mathrm{s}^{-1}), \tag{1}$$

where c is the speed of sound (340 m/s), V is the volume of the room in m^3, and t is elapsed time in seconds after arrival of the direct sound at the receiver position.

Using Schreiber's result for the maximum mean impulse density and the volume of the room, Eq. (1) can be used to calculate the time of onset of the subjectively diffuse part of the sound field. Note that this assumes that a subjectively diffuse sound field is equivalent to a subjectively continuous sound field.

The dimensions and acoustic properties of the listening room of The Acoustics Laboratory was used as basis for the simulation. This room is built in accordance with IEC 268-13 (1985) and is therefore believed to be generally representative of a domestic listening room. The dimensions of the room and the positions of the loudspeaker and listener which have been modeled are shown in Fig. 1.

The room has a volume of 112.8 m^3, which together with a mean impulse density of 2000 s^{-1} is used to solve Eq. (1). This results in an onset time of 21.4 ms relative to the direct sound for the subjectively diffuse part of the sound field.

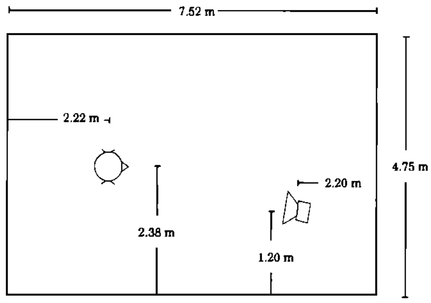

FIG. 1. Plan view of the listening room that defined the input parameters to the image model. The height of the room is 3.16 m with a suspended ceiling at 2.8 m. The center of the loudspeaker and the head of the listener are both 1 m above the floor. The loudspeaker was aimed towards the listener and the distance between the listener and the loudspeaker was 3.1 m.

The sound field, produced by the arrangement shown in Fig. 1, can then be divided into three components:

(1) the direct sound,
(2) individual reflections arriving earlier than about 21–22 ms, relative to arrival of the direct sound, and
(3) the reverberant field or reflections arriving later than 21–22 ms, relative to arrival of the direct sound.

The total number of reflections with delays in the range 0–22 ms is too large for a practical implementation if each reflection is to be represented by an individual loudspeaker in the anechoic chamber. To reduce the number of loudspeakers the following procedure was employed: (1) Only reflections with a level above −20 dB, *re:* the direct sound were modeled and (2) several reflections were represented by the same loudspeaker if their angles of incidence were not subjectively different.

The minimum level of −20 dB was based on threshold values for a single reflection in combination with the direct sound (no other reflections or a reverberant field are present). This situation was assumed to represent the most sensitive situation and to result in threshold values below those exposed for the present situation.

Blauert (1983) provides information on the necessary separation between sound sources to ensure perception as spatially separated, used to represent several reflections by one loudspeaker. Detailed information on the process of reducing the number of loudspeakers is given in Bech (1990).

Thus the implemented simulation included the direct sound, 17 individual reflections, and the reverberant field. The following sections discuss the practical implementation of the three parts of the setup.

A. Implementation of the direct sound and individual reflections

Delay and attenuation due to distance of the direct sound and individual reflections are calculated using the image source theory, implemented as a computer program by KEF Audio Ltd. The directivity characteristics of the *original* loudspeaker were modeled as a cardioid, independent of frequency. The absorption coefficients of the room surfaces were modeled as independent of frequency with the following values: ceiling=0.05, floor=0.3, and walls=0.44. These values result in a calculated reverberation time of 0.4 s, independent of frequency.

The calculated levels of the individual reflections are hereafter referred to as the natural levels, as they represent the levels that the reflections would have in a real room with specified properties.

For a general discussion of the image source principle, see, e.g., Cremer and Müller (1982) and Berman (1975) for use in a rectangular room.

The loudspeakers employed for the simulation setup consisted of a 110-mm, two-way coincident-source drive unit, mounted in a rigid plastic sphere of 280-mm diameter. All loudspeakers were individually calibrated to have a frequency response of 200 Hz–13 kHz±0.5 dB, and matched in sensitivity (200 Hz–13 kHz) within ±0.25 dB. The loudspeaker system is described in detail in Fincham *et al.* (1989). The setup was positioned in the large (1000 m^3) anechoic chamber of The Acoustics Laboratory, and all loudspeakers were located, with correct azimuth and elevation, on the surface of an imaginary sphere of 3-m radius centered on the listening position. The supporting structures for the loudspeakers and the subject were treated with sound absorbing mineral wool to reduce the residual reflections to a level of at least 21 dB below the direct sound, for the frequency range 80 Hz–20 kHz.

B. Implementation of the reverberant field

The creation of a subjectively diffuse sound field in an anechoic chamber has been studied by Meyer *et al.* (1965), Damaske (1967/68), Wagener (1971), Damaske and Ando (1972), and Ando and Kurihara (1986). The general conclusion is that a number of loudspeakers fed with uncorrelated signals, and positioned in the horizontal plane of an imaginary sphere, will provide a listener at the center of the sphere with a subjectively diffuse sound field. The number of loudspeakers necessary depends on the characteristics of the sound field being simulated.

A series of pilot experiments (see Bech, 1990) showed that the present simulation required six loudspeakers, positioned in the equatorial plane of the imaginary sphere described above. Signals for the reverberant field were based on a commercially available reverberation unit (Lexicon PCM70) which provides two sufficiently uncorrelated outputs.[1] The reverberation time measured in the setup was in the range 0.4–0.5 s for the frequency range 125 Hz–5000 Hz. The level of the reverberant part of the sound field, relative to the direct sound and early reflections, was calculated assuming a reverberation time of 0.4 s at 1 kHz, and an exponential decay. The system was calibrated using broadband pink noise.

A block diagram of the complete system is shown in Fig. 2. The positions of all the loudspeakers, and the delay and attenuation of all signals representing individual images and reverberation channels, are given in Table I. In the following, individual reflections will be identified either by delay, *re:* the direct sound, or the number given in Table I.

The design criteria and implementation of the reverberant field is further discussed in Bech (1990).

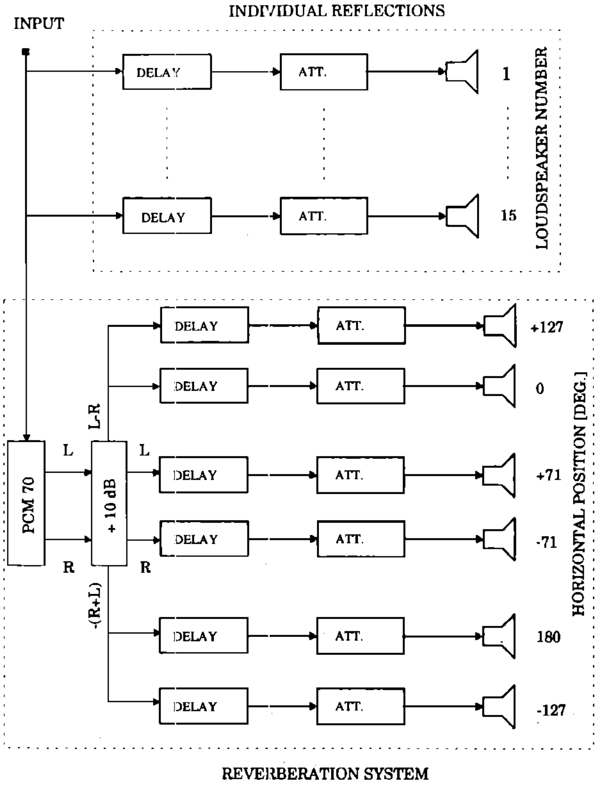

FIG. 2. Block diagram of the complete experimental setup. Note that the reflection loudspeakers can represent more than one image, cf. Table I. Note that 0° corresponds to the front angular reference for the subject and positive angles are to the left-hand side.

C. Subject positioning and calibration procedures

The listener's ears must be located close to the specified listening position. A motorized adjustment mechanism was built into the chair supporting the subject to position the listener's ears in the correct position, as observed through a fixed video camera. An additional video camera was used for monitoring and recording the listener's head movements during listening. A curtain prevented the listener from seeing the simulation setup, while a single red LED was used to define the front angular reference. Listeners were free to move their heads, but were instructed to focus attention on the LED.

A computer-controlled calibration system was built to ensure that the system functioned to specifications. The frequency response of every image was measured each day and checked against a reference, before and after measurements. All loudspeakers were checked for obvious audible distortion by means of a sinusoidal sweep.

The reproduction level, measured at the listening position with a single microphone, was 66 dB SPL for the noise stimulus, and approximately 50 dB SPL (time weighting fast) for the speech stimulus. The level of the speech stimulus was set to correspond to normal conversational level at a 3-m distance in a listening room. The overall background noise level for all elements of the system was 27 dB SPL (time weighting fast). The one-third octave background noise level was constant, at ±2 dB for the frequency range 20 Hz–20 kHz.

TABLE I. Position of loudspeakers and delay and attenuation of the signals to the loudspeakers for primary loudspeaker, images, and reverberation channels included in the simulation setup. The last surface of the reflection path is also given. All angles and wall references are relative to the listening position. The left-hand side of the subject defines positive angles.

Delay (ms)	Att. (dB)	Azimuth	Elevation	Reflection No.	Surface
0.00	0.0	−22°	0°	···	primary lsp
1.64	3.6	−25°	−28°	1	floor
4.16	9.2	−50°	−2°	2	right wall
4.48	5.0	−25°	48.2°	3	ceiling
5.36	11.6	−53°	−28°	4	floor
7.60	11.8	−50°	48°	5	ceiling
9.20	10.0	−25°	48.2°	6	ceiling
9.20	10.0	−25°	−56°	7	floor
9.94	9.7	65°	0°	8	left wall
10.80	11.8	65°	−14°	9	left wall
11.64	15.5	−53°	−56°	10	floor
11.64	15.5	−50°	48°	11	ceiling
12.50	11.5	65°	30°	12	left wall
12.70	9.9	−170°	0°	13	back wall
13.46	11.9	−170°	−15°	14	back wall
14.42	14.3	−25°	−56°	15	floor
14.80	14.6	−154°	0°	16	back wall
14.98	11.3	−170°	33°	17	back wall
22.00	0.5	71°	0°	···	rev. syst.
22.00	0.5	−71°	0°	···	rev. syst.
23.00	7.5	127°	0°	···	rev. syst.
24.00	7.5	−127°	0°	···	rev. syst.
25.00	8.5	180°	0°	···	rev. syst.
26.00	0.5	0°	0°	···	rev. syst.

II. STIMULI

Previously published results have shown that the type of signal has a strong influence on the threshold of detection for a single reflection in combination with the direct sound [see Olive and Toole (1989) for a review]. Two types of sound have typically been used, discontinuous sound like clicks, castanets and speech, and continuous sounds like noise (pink or white) or music recorded under fairly reverberant conditions. It was decided to use pink noise and speech as representatives of the continuous and discontinuous sounds, respectively.

The noise signal was a 1-s sample of broadband (20 Hz–20 kHz) pink noise, and the speech signal was a 3.8-s sample of male speech. The spoken text was a recording, made in the large anechoic chamber of The Acoustics Laboratory, of an excerpt of the text used for the standardized Danish speech material for audiometric purposes (Elberling *et al.*, 1989). The time structure and spectrum of the chosen speech sample resemble average Danish speech.

The noise and speech samples were transferred to a hard disk based editing system (Ariel DSP-16 SDI system) and played back via 16-bit D/A at a sampling rate of 50 kHz with low-pass filtering at 20 kHz. The rise and fall time of the envelope of the noise signals is 5 ms following a linear function. The same sample of the noise and speech signal was used in all presentations. For further details of the recordings made for the project, see Hansen and Munch (1991).

TABLE II. Overview of the experiments which are discussed in this paper. The reflection numbers refer to Table I. Note that the abbreviation ISI stands for interstimulus interval.

Exp. No.	Description of experiment	Signal	No. of subjects	Results in Fig.
I	measurement of TD for reflections Nos.			
	*1–17	noise	8	3
	*1,2,3,6,7,8,13,17	speech	4	3
II	measurement of TD without the reverberant field for reflections Nos.			
	*1,2,8,11, and 16	noise	8	4
III	measurement of TD for ref. Nos. 1 and 17 with a 10-dB decrease in total SPL	noise	8	not shown
IV	measurement of TD for ref. 1 with ISI increased by 0.5 to 1 s.	noise	8	not shown
V	measurement of jnd for reflections Nos.			
	*1–17	noise	8	5
	*1,2,3,6,7,8,13,17	speech	8	5

III. SUBJECTS

Eight subjects (five male and three female) participated in the experiments. All had hearing thresholds within 5 (male) and 15 dB (female), *re:* ISO 389 (1985) at all audiometric frequencies. The number of subjects participating in each experiment is given in Table II. The subjects were selected from students at the laboratory, on the condition they were available for the full project period (five years). They were paid an hourly rate for participating in the experiments. None of the students had any previous experience in psychoacoustic experiments.

Each subject received a minimum of 5000 trials (noise) and 1200 trials (speech) as a part of a training program conducted before the main experiments. A general discussion of training and an examination of the learning process in the present series of experiments are given in Bech (1993a).

IV. GENERAL PROCEDURE

The subject's task in all experiments was to detect a change in timbre of a pink noise stimulus or a speech stimulus. The interpretation of timbre [see, e.g., ANSI S1.1 (1960)] was discussed with the subjects during the training period and whenever they wished to do so during the experimental period. See the detailed instruction in Bech (1994b).

For each reflection in the group of 17 early reflections, two psychoacoustic quantities were determined: (a) threshold of detection and (b) just-noticeable difference corresponding to questions (1) and (2), respectively, as asked in the introduction. An adaptive (staircase) two-alternative forced-choice procedure was used.[2] The standard and the comparison stimuli for the two situations are defined as follows:

(1) Threshold of detection (TD)—The *standard* is the complete sound field simulating a loudspeaker in the listening room, except that the reflection under investigation is absent (i.e., attenuated 100 dB, *re:* direct sound). The *comparison* stimulus was formed by adding a variable level of the reflection under investigation to the standard.
(2) Just-noticeable difference (jnd)—The *standard* was the complete sound field simulating a loudspeaker in the listening room. The *comparison* stimulus is derived from the standard by a variable increase in the level of the reflection under investigation.

In the TD experiments, the initial level of the reflection under investigation was equal to the level of the direct sound. For the jnd experiments, the initial level was a 10-dB increase of the level of the reflection in the standard.

The level of the reflection under investigation was varied adaptively (two down/one up) to estimate that level, which would produce 70.7% correct responses (Levitt, 1971). The step size initially 4 dB, was reduced to 2 (TD experiments) or 1 dB (jnd experiments) after three reversals. Typically, 10–15 reversals would occur during each 50-trial block. For each block, the threshold was estimated as the average of the levels at the midpoints of runs 4, 6, 8, etc. The reported TD or jnd are averages across subjects for eight (noise) or six (speech) 50-trial block estimates per subject.

The comparison stimulus was present either in the first or second observation period with equal probability. The other period contained the standard. The two observation periods were separated by an interstimulus interval of 0.5 s.

The subject responded (to indicate the comparison stimulus) by pressing a button on a small keyboard placed at the subject's right-hand arm rest. This keyboard included a small screen on which events within a trial and feedback (right/wrong response) were displayed. A feedback was also given acoustically after each response (two or four beeps).

The experiments were conducted in such a way that each subject participated in one session per day. A session concluded eight blocks of 50 trials (noise) or six blocks of 50 trials (speech), taking about 2 h. One reflection was examined per subject per session. Two subjects participated in each session, and while one ran two blocks in succession, the other rested. The two blocks, run in succession, were separated by a 5 to 10-min rest period. After two blocks, the subjects interchanged. The order of presentation of the reflections was such that order effects were balanced in experiments I and V. A random order of presentation was used for the other experiments. An overview of all experiments is given in Table II.

V. RESULTS AND DISCUSSION

A. Threshold of detection for individual reflections

The purpose of these experiments (Nos. I–IV in Table II) was to measure the threshold of detection for individual reflections under different conditions, and to compare the measured values with the natural level of the reflection in a standard listening room. The natural level was calculated using the image model as described in Sec. I. The experiments are related to question one in the Introduction. Note that the measured thresholds for reflections Nos. 6, 7, 10, and 11 are not discussed here or in any of the following sections, as they are treated separately under the general discussion.

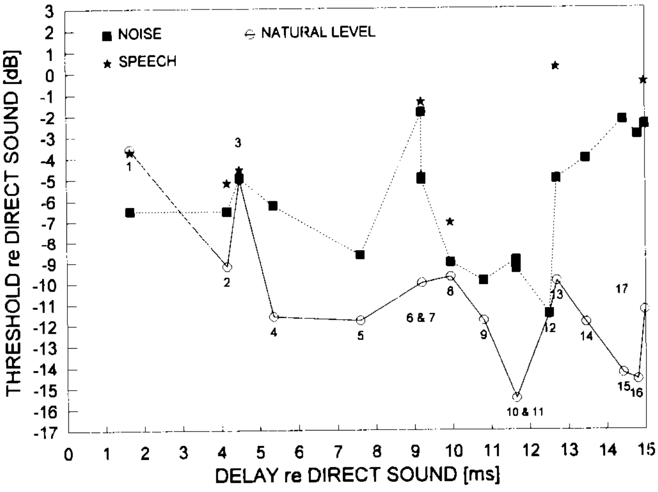

FIG. 3. Threshold of detection for individual reflections for the noise and speech stimulus (experiment I). The natural levels of the individual reflections based on an image model are also shown as are the individual reflection numbers according to Table I. Confidence (95%) intervals are ± 0.8 (noise) and ± 1.1 dB (speech). The confidence intervals are based on the variance between blocks and the mean values are based on eight subjects and 400 (noise) or 300 trials (speech) per subject.

1. Comparison of natural levels and measured thresholds of individual reflection

The thresholds of detection for noise and speech signals are shown in Fig. 3, together with the natural levels.

The general tendency is for the natural levels to be lower than the measured TDs. The only exceptions are for reflections 1 (floor), 3 (ceiling), 8 (left wall), and 12 (left wall) for the noise signal, where the natural levels are either higher, or not statistically significantly lower, than the TDs. This suggests that only these reflections will contribute to the timbre on an individual basis.

Generally, the use of the speech signal is seen to increase the threshold, most notably for reflections 1 and 3, to such a degree that only reflections 1 and 3 have natural levels which are above or at the TDs.

The increase in threshold values for the speech signal is probably a result of uncertainty introduced by the nonstationary character of the signal. Green (1988) has shown that any uncertainty in the signal parameters will cause detection performance to decrease. Differences between the threshold values for the noise and speech signals in the present series of experiments are smaller than those found by, for example, Olive and Toole (1989). However, this could be explained by the fact that the same 3.8-s speech segment has been used for all the speech experiments, leading to a lower degree of uncertainty than if different speech segments had been used. The same noise segment was also used in all the present noise experiments, so a similar reduction in uncertainty would be expected for the noise signal. However, Buus (1990) has shown that the improvement in detection perfor-

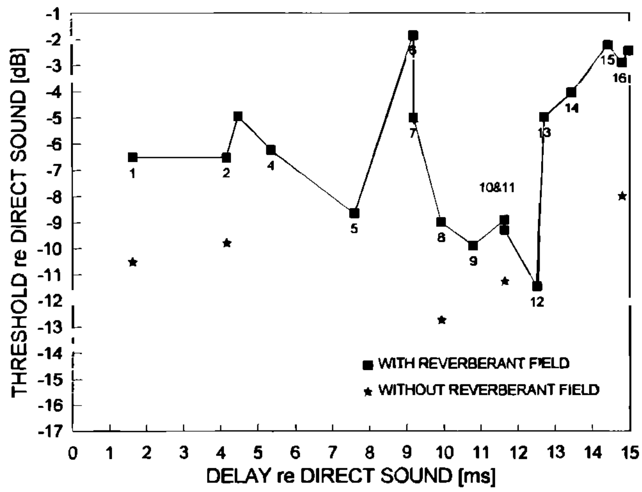

FIG. 4. Threshold of detection for individual reflections for the noise stimulus with (exp. I) and without (exp. II) the reverberant field included in the total sound field. Confidence intervals (95%) are ±0.8 (exp. I) and ±1 dB (exp. II). The confidence intervals are based on the variance between blocks and the mean are based on eight subjects and 400 trials per subject. The individual reflection numbers are shown according to Table I.

mance by using a "frozen" noise signal (wideband) is negligible compared with the use of random noise signal in every presentation.

Thus these preliminary results suggest that only the first-order floor (ref. No. 1) and ceiling (ref. No. 3) reflections will be potentially audible as individual reflections, in the total sound field produced by a speech program and a loudspeaker in a listening room. The results further indicate that if a signal with higher certainly than speech, such as broadband noise, is used, other reflections in addition to the first-order floor and ceiling reflections might become audible.

2. *The influence of the reverberant field*

To examine the influence of the reverberant field, TD values were measured for selected reflections without the reverberant field. The results, shown in Fig. 4, show that removal of the reverberant field causes the threshold values to decrease by 2–5 dB. The tendency as well as the magnitude of the change are in agreement with the results of Schubert (1969) and Olive and Toole (1989).

These results suggest that the contribution of individual reflections to the timbre of the sound field will increase with decreasing level of the reverberant field. It follows that the level of individual reflections should be considered more carefully in, for example, control rooms which typically have lower reverberation times compared to domestic rooms.

The reason for the decrease in threshold values when the reverberant field is absent can only be speculative at present. However, the following might throw some light on the process.

It is generally agreed that the principal dimension in the timbre space is correlated with a physical parameter based on spectral information (Pols, 1970; Plomp, 1970; Bismark, 1974a,b; Grey, 1975; de Bruijn, 1978). Plomp (1970) also showed that predictions of timbre from a model based on the spectral differences between two stimuli correspond to perceived differences in timbre.

The addition of a single reflection to the standard will introduce spectral changes corresponding to the effects of a comb filter. The magnitude of the changes depends on the relative level of the reflection. However, changes will be smaller if energy, uncorrelated with the direct sound and the individual reflections, is added to the sound field in the form of a reverberant field.[3]

It is therefore suggested that the reverberant field will act as a masker, reducing the spectral differences between the two stimuli, and result in an increase of the detection threshold for the individual reflections.

3. *Influence of reproduction level and interstimulus interval*

As the removal of the reverberant field (exp. II) lowers the overall reproduction level, it must be verified that the differences in threshold levels, shown in Fig. 4, are not due to the difference in reproduction levels. Parts of experiment I were therefore repeated with a 10-dB reduction in reproduction level and the TDs were measured for reflections 1 and 17.

The results showed that the thresholds did not change significantly at the lower reproduction level. This agrees with results by Bilsen (1968) and Schubert (1969) who showed that the TD for a single reflection in combination with the direct sound is independent of reproduction level, if the absolute level of the reflection is above the hearing threshold.

Thus it can be concluded that the decrease in reproduction level, caused by removal of the reverberant field, is not likely to have influenced the threshold values obtained in experiment II.

The influence of the length of the interstimulus interval (ISI) was examined because, with a reverberant time of 400 ms, the possibility that the "tail" of the first stimulus of a pair could influence the second with an ISI of only 500 ms could not be excluded. The TD was thus measured for reflection 1 using noise and an ISI of 1 s. The thresholds for ISIs of 500 ms and 1 s were not significantly different, and it was concluded that the use of an ISI of 500 ms had not influenced the thresholds obtained.

B. Just-noticeable difference in level for individual reflections

The purpose of these experiments was to measure the just-noticeable difference (jnd) in the level of individual reflections in the complete sound field. This is related to question (2) in the Introduction.

The changes in the level of individual reflections could, for example, be a result of changing the listening position relative to the loudspeaker. The results of the jnd experiments would thus provide information on which, and how much, individual reflections should be changed, in order to maintain a constant timbral character of the reproduced sound.

The detection thresholds which were discussed in the previous sections would seem to make the measurement of

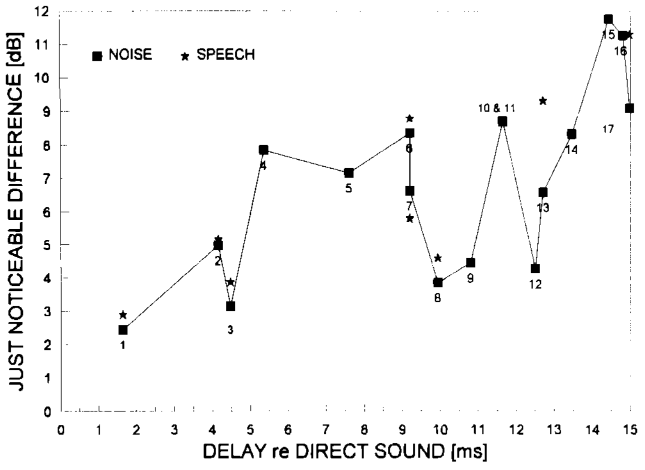

FIG. 5. Just-noticeable difference for an increase in level for individual reflections for the noise and speech signal. Confidence intervals (95%) are ±0.45 (noise) and ±0.7 db (speech). The confidence intervals are based on the variance between blocks and the mean values are based on four subjects and 400 (noise) or 300 trials per subject. The individual reflection numbers are shown according to Table I.

jnd's superfluous: The majority of the TDs are above the natural level predicted by the image model. The apparent consequence is that the negative jnd's (a decrease in the natural level) could not be measured and that the positive jnd's would be simply the difference between the natural level and the detection threshold.

Pilot experiments with a few reflections showed that the negative jnd's could not be measured, as reflections could be attenuated by 100 dB without any subjective effects. However, it was also found that the positive jnd's were significantly higher than the difference between the natural level and the detection threshold. It was therefore decided to measure the positive jnd's for all reflections using the noise signal, and for the eight strongest reflections using the speech signal.

1. Results

The positive jnd values for the noise and speech signal are shown in Fig. 5. The jnd's are seen to divide the reflections into two groups: one group includes reflections 1, 2, 3, 8, 9, and 12 and the other contains the remaining reflections. A similar grouping was found in experiment I where reflections 1, 3, 8, and 12 have natural levels that are either higher or similar to the TDs and the remaining reflections have natural levels below the TDs.

Thus both the TD and jnd results suggest that reflections from the floor (ref. No. 1), the ceiling (No. 3), and the left wall (Nos. 8, 9, and 12) will influence the timbre of the sound field to a higher degree than any of the other reflections investigated.

Statistically, the jnd's for the two signals are not significantly different except for reflections Nos. 13 and 17. While this is the same tendency as observed for the TD results, the influence of signal type is generally smaller for the jnd results than for the TD results.

VI. GENERAL DISCUSSION

The general discussion will include the following points:

(1) the detection cues for TD and jnd,
(2) results for reflections 6, 7, 10, and 11,
(3) training effects, and
(4) generality of the results.

A. Detection cues at TD and jnd

Throughout this paper it has been assumed that the subjects followed the instructions and used timbral differences to established the TD or jnd. However, at least three different attributes were available in the course of the adaptive TAFC procedure: At the initial level of the reflection in both TD and jnd experiments, the standard and the comparison stimuli could differ in auditory position of the main sound source (image shift), loudness, and timbre. When the level of the reflection was reduced, the image shift was the first cue to disappear. For a further reduction in level, the loudness cue disappeared, and then finally at the threshold, the timbral cue disappeared. When the timbral cue disappeared, there was no perceptual difference between the two sound fields.

The three cues were discussed carefully with the subjects and none of them reported any difficulties in distinguishing between image shifts and timbre differences. This has later been confirmed in another experiment where TDs were measured in separate experiments using timbre and image shift as detection cues, respectively, for the same subjects. The results showed that the two sets of TDs were significantly different (see Bech, 1994a). Olive and Toole (1989) also showed that an image shift based threshold is significantly different from one based on timbre detection.

To examine the possible use of loudness differences as cues, the SPL differences between the standard and the comparison stimulus with the reflection at TD for noise (exp. I) were measured. The measuring procedure ensured that objective measuring errors would not influence results. The measured SPL differences were below 0.25 dB for reflections with delays smaller than 13 ms, and in the range of 0.3–0.55 dB for longer delays. Miller (1947), Zwicker and Feldtkeller (1967), Houtsma *et al.* (1980), and Schorer (1989) have shown that the jnd for a level difference between two broadband noise signals is in the range 0.5–0.65 dB. The objective measurements thus confirm that loudness was not been used as a detection cue at the threshold, for reflections with delays less than 13 ms. However, they also indicate that loudness based detection could have been used, as part or as sole criterion, for reflections with delays above 13 ms.

B. Results for reflections 6, 7, 10, and 11

The results (TD and jnd) for reflections 6,7, 10, and 11 have not been discussed previously because they deviate from the results for the other reflections. The general trend for reflections 6, 7, 10, and 11 are summarized in Fig. 6, which shows the results for TD and jnd+ for all reflections and the noise signal. The jnd+ results are the natural levels increased by the positive jnd values shown in Fig. 5.

The TD values for reflections 6 and 7 and 10 and 11, respectively, are higher than the TDs for reflections with similar delay times, that is, reflections 5 and 8 and 9 and 12, respectively. The two groups of reflections are special, as both reflections in the group have the same delay times relative to the direct sound. To test if this has any influence on the threshold values for individual reflections, the TDs were measured for reflections 6 and 7 individually, without the other reflection present. The results showed that the TDs decreased by 4.3 and 3.5 dB for reflections 6 and 7, respectively, when the other reflection was not present. This strongly suggests that the presence of multiple reflections with similar delay times results in a masking effect for individual reflections.

This masking effect could also be the reason for the general difference between the TD and jnd+ results, because the standard in the jnd experiment includes the reflection of interest at the natural level, and at an increased level in the comparison stimulus. This is similar to the original TD experiment for reflections 6 and 7, as the standard includes one reflection and the comparison stimulus contains the other. The observed differences between the TD and jnd+ results are also of similar magnitude to the differences between the TDs for reflections 6 and 7, with and without the other reflection present.

A simple explanation of this masking effect could be the following: The spectrum of the standard will contain comb filter characteristics that are determined by the relative levels and delays of the individual reflections present in the standard. In the TD experiments, spectral differences between the standard and the comparison stimuli will be those *new* comb filter characteristics that are generated by adding the investigated reflection to the standard. However, in the jnd experiments the spectral differences will be a *change* of an already existing comb filter pattern, as the investigated reflection is present in both stimuli. Thus to produce the same spectral difference between the two stimuli, the relative level of the reflection needs to be higher in the jnd experiment. Note that it is assumed that the timbral difference between the two stimuli is generated by the spectral difference between the two. This assumption is also used for the discussion of the influence of the reverberant field (see Sec. V A 2).

C. Effects of training

The subjects were exposed to an extensive training program before they participated in the main experiments (see Bech, 1993a). This initial training was to ensure that their thresholds had reached an asymptotic level. The initial training experiments were repeated at regular intervals throughout the period of experimentation to ensure that the subjects maintained their performance.

D. Generality of the results

The generality of the results which have been presented in this paper depend on a large number of factors. The most important of these will be discussed in the following.

The first factor is the ability of the simulation setup to convey an impression to the subjects perceptually similar to that of a real room.

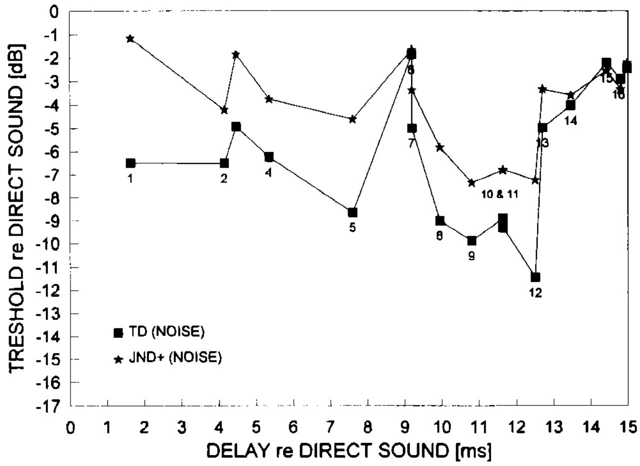

FIG. 6. Comparison of threshold of detection values from experiment I and jnd+ values from experiment V for the noise signal. Note that the jnd+ values are the levels calculated by the image model and increased by jnd. The individual reflection numbers are shown according to Table I.

The electroacoustic simulation principle has been used most noticeably in the area of concert hall acoustics where it is generally accepted that the realism of the simulation, if done properly, can be very high [see Cremer and Müller (1982) for a review]. The main problem with any simulation technique is to verify the degree of realism of the simulation. The present setup has been evaluated on an informal basis by inviting researchers with experience in electroacoustic simulation to listen to the setup and evaluate quality. It has been the general impression of the visitors that the setup so far represents the most accurate simulation of the conditions in a listening room. This means that the generality of the results should be as good and possibly better than has previously been obtained with simulation setups.

A second factor is the accuracy of the parameters used to define the simulation. The main parameters are the acoustic characteristics of the room and the loudspeaker, and the positions of the listener and the loudspeaker.

The listening room which has formed the basis for the simulation is designed according to the IEC 268-13 (1985) recommendation. This ensures that the room dimensions and reverberation time as a function frequency are representative of the listening rooms of average listeners. The modeled absorption coefficients of the room surfaces have not included any frequency dependence, and the directivity of the loudspeaker has been modeled as a cardioid independent of frequency. This must be assumed to limit the realism of the simulation. The effect of including frequency-dependent absorption and the directivity characteristics of a real loudspeaker will be discussed in a forthcoming paper.

A third factor is the loudspeaker position which is known to have an influence on the perceived sound quality. As the degree of influence was not known at the start of the experiments, it was decided to conduct a series of listening tests in real rooms to examine this effect further. The result of this investigation is reported in Bech (1993b).

The above discussion thus suggests that care should be taken when generalizating from the results presented in this paper.

VII. SUMMARY OF FINDINGS

This section contains a summary of the main findings. They are all based on an electroacoustic simulation of the right-hand loudspeaker of a stereophonic setup, positioned in a small room. The validity of the results is discussed in Sec. VI D.

A. Threshold of detection experiments

The results show that only the first-order ceiling and floor reflections are likely to contribute individually to the timbre of a speech signal. For a noise signal additional reflections, from the wall to the left of the listener, will individually contribute to the timbre.

The threshold of detection for all reflections depends on the level of the reverberant field. If the reverberant field is removed, thresholds will decrease by 2–5 dB.

B. Just-noticeable-difference experiments

The results show that an increase in the level of individual reflections is most likely to be audible for the first-order floor and ceiling reflections, and for reflections from the wall to the left of the listener. This applies to both the speech and noise signal.

C. General

The results show that the TD and jnd values for individual reflections in a group will be higher if multiple reflections in the group have the same delay relative to the direct sound.

ACKNOWLEDGMENTS

The project team included the following on a permanent basis: L. R. Fincham, Dr. R. H. Small, A. Jones, R. I. Harris, and R. Wilson, all from KEF Audio Ltd., V. Hansen, G. Munch, and K. B. Christensen, all from Bang and Olufsen Ltd., and K. Rasmussen, O. Juhl Pedersen, N. Farver, and S. Bech, all from The Acoustics Laboratory. The entire group of subjects and experimenters are thanked for their enthusiasm and support throughout the project. Finally, Dr. R. H. Small, O. J. Pedersen, S. K. Pramanik, and G. Munch are thanked for many useful comments on earlier versions of this manuscript and Dr. G. Naylor for comments on earlier versions of this paper and for many hours of discussion in the course of the project. The author would also like to thank Professor D. M. Green, Dr. H. Lehnert, Dr. F. Giron, and the reviewers for very helpful comments.

[1]The discussion in Bech (1990) shows that it is the maximum absolute value of the normalized cross-correlation function of the two ear-input signals, within a correlator time shift (τ) window of ± 1 ms, that is important for the perceived diffuseness of the sound field. A value of approximately 0.15 was found to be sufficient and the present reverberation system results in a value of 0.17.

[2]This section attempts to explain the adaptive two-alternative forced-choice (TAFC) procedure that was employed for the experiments. The explanation will closely follow that of Levitt (1978). An adaptive procedure is characterized by a stimulus level in any one trial that is determined by the response and stimulus levels used in the preceding trials.

One trial includes two observation intervals separated by an interstimulus interval. After the second observation interval follows an answering period. In one of the observation intervals (random choice) the standard is presented and the other contains the comparison stimulus. The subject's task is to indicate (during the answer period) which of the two observation intervals contained the comparison stimulus. After the answer interval a feedback (right/wrong response) is given.

The standard stimulus is constant throughout a block of trials, whereas the comparison stimulus contains the parameter that is varied. A classical example of a standard stimulus is a noise burst and the corresponding comparison stimulus could be a noise burst plus a tone. The level of the tone is varied according to the rules of the adaptive procedure and the subject's task is to detect the presence of a tone in the noise. The level of the parameter (e.g., the tone) of interest is termed the stimulus level.

The threshold value is based on a series of trials and the stimulus level in the first trial is termed the initial value. This value is often chosen well (10–15 dB) above the expected threshold level. The next stimulus value depends on the response, the adaptive rule, and the step size. The adaptive rule prescribes how many consecutive positive (i.e., the subject correctly identifies the comparison stimulus) or negative responses that are needed before the stimulus level is decreased or increased, respectively. An often used rule is the so-called up–down where a single positive or negative response leads to a decrease or increase , respectively, in the stimulus level. The up–down rule results in a threshold level which on the average leads to 50% positive responses. The rule used for the present experiments is the so-called two-down/one-up where two positive responses lead to a decrease in the stimulus level and one negative leads to an increase. This rule produces a threshold level that on the average will produce 70.7% correct responses. The stimulus level changes in steps determined by the chosen step size. The step size is often changed from a high to a low value (typically from 4 to 2 dB) after a certain number of trials.

Other terms often used for describing the procedure are "run," "reversal," and "midpoint." One run is defined as a sequence of trials where the changes in stimulus level were all in one direction. A reversal is the point (the trial) where a change is made in the direction of stimulus adjustment. A midpoint level is the mean value of the levels at two succeeding reversals. The threshold level is often calculated as the average of a number of midpoint levels (e.g., for runs 2, 4, 6, 8, etc.). Only every second run is used as a certain degree of correlation exists between succeeding runs. A reliable threshold estimate must be based on a certain number of midpoint levels which requires a certain number of trials, and a typical value is 50 trials.

[3]The reverberant field will generally be uncorrelated with individual reflections in a real second field produced by a noise signal [for a further discussion of this see Sec. IV 1 3 5 in Cremer and Müller (1982)]. However, in the simulation setup the reverberant field was produced by a commercial reverberation unit that had time varying elements in the algorithm and the specific details of the process were not available. Thus it is difficult to estimate the degree of correlation between the reverberant field and the direct sound and early reflections. However, a reasonable assumption, based on the available measurements and the subjective effect it produced, is that the correlations are of the same magnitude as for a real sound field.

ANSI S1.1-1960, "Acoustical terminology" (American National Standards Institute, New York).

Ando, Y., and Kurihara, Y. (**1986**). "Nonlinear response in evaluating the subjective diffuseness of sound fields," J. Acoust. Soc. Am. **80**, 833–836.

Atal, B. S., Schroeder, M. R., and Kuttruff, K. H. (**1962**). "Perception of coloration in filtered Gaussian noise—Short-time spectral analysis by the ear," Paper H31, Fourth International Congress on Acoustics, Copenhagen, 21–28 August.

Bech, S. (**1989**). "Audibility of single reflections with short delay times," 13th International Congress on Acoustics, Belgrade, Vol. 1, pp. 447–450.

Bech, S. (**1990**). "Electroacoustic simulation of listening room acoustics; Psychoacoustic design criteria," 89th Convention of the Audio Engineering Society, Los Angeles, USA, 21–25 Sept., Preprint No. 2989.

Bech, S. (**1993a**). "Training of subjects for auditory experiments," Acta Acoust. **1**, 89–99.

Bech, S. (**1993b**). "Perception of timbre of reproduced sound in small rooms: The influence of the room and the loudspeaker position," J. Audio Eng. Soc. **42**, 999–1007.

Bech, S. (**1994a**). "Auditory localization of reproduced sound in small rooms."

Bech, S. (**1994b**). "Perception of reproduced sound: Audibility of individual reflections in a complete sound field," 96th Convention of The Audio

Engineering Society, Amsterdam, The Netherlands, 26 February–1 March, preprint No. 3849.

Berman, J. M. (**1975**). "Behavior of sound in a bounded space," J. Acoust. Soc. Am. **57**, 1275–1291.

Bilsen, F. A. (**1968**). "On the interaction of a sound with its repetitions," Ph.D. dissertation, Uitgeverij Waltman, Delft.

Bismark, G. von (**1974a**). "Timbre of steady sound: A factoral investigation of its verbal attributes," Acustica **30**, 146–159.

Bismark, G. von (**1974b**). "Sharpness as an attribute of the timbre of steady sounds," Acustica **30**, 159–172.

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, London).

Bruijn, A. de (**1978**). "Timbre classification of complex tones," Acustica **40**, 108–114.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

Cremer, L., and Müller, H. A. (**1982**). *Principles and Applications of Room Acoustics*, translated by T. J. Schultz (Applied Science, New York), Vol. 1, pp. 17–19.

Damaske, P. (**1967/68**). "Subjektive Untersuchungen von Schallfeldern," Acustica **19**, 199–123.

Damaske, P., and Ando, Y. (**1972**). "Interaural crosscorrelation for multichannel loudspeaker reproduction," Acustica **27**, 232–238.

Elberling, C., Ludvigsen, C., and Lyregaard, P. E. (**1989**). "Dantale: A new Danish speech material," Scan. Audiol. **18**, 169–175.

Fincham, L. R., Jones, A., and Small, R. H. (**1989**). "The influence of room acoustics on reproduced sound, part 2: Design of wideband coincident-source loudspeakers," 87th Convention of the Audio Engineering Society, New York, USA, 18–21 Oct., preprint No. 2877.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).

Grey, J. M. (**1975**). "An exploration of musical timbre," Ph.D. thesis, Department of Music, Report No. STAN-M-2, Stanford University.

Hansen, V., and Munch, G. (**1991**). "Making recordings for simulation tests in the Archimedes project," J. Audio Eng. Soc. **39**, 768–774.

Houtsma, A. J. H., Durlach, N. I., and Braida, L. D. (**1980**). "Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching," J. Acoust. Soc. Am. **68**, 807–813.

IEC (**1985**). IEC Recommendation 268-13, "Sound system equipment, Part 13: Listening tests on loudspeakers" (International Electrotechnical Commission, Geneva).

ISO (**1985**). ISO No. 389–1985 (E), "Standard reference zero for the calibration of pure tone air conduction audiometers," (ISO, Geneva).

Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Levitt, H. (**1978**). in "Sensori-neural hearing impairment and hearing aids," Scand. Audiol. Suppl. **6**, 241–291.

Meyer, E., Burgtorf, W., and Damaske, P. (**1965**). "Eine Apparatur zur elektroakustischen Nachbildung von Schallfeldern. Subjektive Hörwirkungen beim Übergang kohärenz—Inkohärenz," Acustica **15**, 339–344.

Miller, G. A. (**1947**). "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," J. Acoust. Soc. Am. **19**, 609–619.

Olive, S. E., and Toole, F. E. (**1989**). "The detection of reflections in typical rooms," J. Audio Eng. Soc. **37**, 539–553.

Plomp, R. (**1970**). "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 397–414.

Pols, L. C. W. (**1970**). "Perceptual space of vowel-like sounds and its correlation with frequency spectrum," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 463–473.

Reichardt, W., and Lehmann, U. (**1978**). "Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit, Erläuterungen des Raumeindrucksmasses R," Acustica **40**, 277–290.

Schorer, E. (**1989**). "Vergleich eben erkennbarer Unterschiede und Variationen der Frequenz und Amplitude von Schallen," Acustica **68**, 183–199.

Schreiber, von L. (**1960**). "Was empfinden wir als gleichförmiges Rauschen?," Freqenz **14**, 399–403.

Schubert, P. (**1969**). "Die wahrnehmbarkeit von Rückwürfen bei Musik," Hochfrequenztechn. Elektroakust. **78**, 230–245.

Somerville, T., Gilford, C. L. S., Spring, N. F., and Negus, R. D. M. (**1966**). "Recent work on the effects of reflectors in concert halls and music studios," J. Sound Vib. **3**, 127–134.

Wagener, B. (**1971**). "Raümliche verteilungen der hörrichtungen in synthetisher schallfeldern," Acustica **25**, 203–219.

Zurek, P. M. (**1979**). "Measurement of binaural echo suppression," J. Acoust. Soc. Am. **66**, 1750–1757.

Zwicker, E., and Feldtkeller, R. (**1967**). *Das Ohr als Nachrichtenemfänger* (Hirzel Verlag, Stuttgart), 2nd ed.